



# Chemometric classification of casework arson samples based on gasoline content



Nikolai A. Sinkov<sup>1</sup>, P. Mark L. Sandercock<sup>2</sup>, James J. Harynuk<sup>1,\*</sup>

<sup>1</sup> Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada

<sup>2</sup> Trace Evidence Services, Royal Canadian Mounted Police, Edmonton, Alberta, Canada

## ARTICLE INFO

### Article history:

Received 30 May 2013

Received in revised form 17 November 2013

Accepted 27 November 2013

Available online 16 December 2013

### Keywords:

Arson  
Chemometrics  
Cluster resolution  
GC–MS  
Gasoline  
Variable selection

## ABSTRACT

Detection and identification of ignitable liquids (ILs) in arson debris is a critical part of arson investigations. The challenge of this task is due to the complex and unpredictable chemical nature of arson debris, which also contains pyrolysis products from the fire. ILs, most commonly gasoline, are complex chemical mixtures containing hundreds of compounds that will be consumed or otherwise weathered by the fire to varying extents depending on factors such as temperature, air flow, the surface on which IL was placed, etc. While methods such as ASTM E-1618 are effective, data interpretation can be a costly bottleneck in the analytical process for some laboratories. In this study, we address this issue through the application of chemometric tools.

Prior to the application of chemometric tools such as PLS-DA and SIMCA, issues of chromatographic alignment and variable selection need to be addressed. Here we use an alignment strategy based on a ladder consisting of perdeuterated n-alkanes. Variable selection and model optimization was automated using a hybrid backward elimination (BE) and forward selection (FS) approach guided by the cluster resolution (CR) metric.

In this work, we demonstrate the automated construction, optimization, and application of chemometric tools to casework arson data. The resulting PLS-DA and SIMCA classification models, trained with 165 training set samples, have provided classification of 55 validation set samples based on gasoline content with 100% specificity and sensitivity.

© 2013 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

Arson is defined as “the act of wilfully and maliciously setting fire to another man’s house, ship, forest, or similar property; or to one’s own, when insured, with intent to defraud the insurers” [1]. Arson damage to residences, businesses, vehicles or other property is but one of the problems; arson also leads to loss of life and feelings of insecurity in the community. Financial costs extend beyond the price of the property damaged, leading to increased insurance rates, costs of fire protection, law enforcement, etc. [2]. Arson tends to be difficult to investigate since much of the evidence is inevitably damaged by the fire [3] as well as by the firefighting efforts, despite best efforts taken to minimize damage to the scene [4]. Important pieces of evidence during a fire investigation include ascertaining the presence of an ignitable liquid (IL) at the scene, as well as the determination of its identity [4].

Due to availability, efficacy, and low cost, petroleum-based accelerants are most often used by arsonists [5]. These ILs may contain hundreds of individual compounds with a specific composition that varies over time and depends on the vendor. Gasoline tends to be the most common IL used in arson since [6,7], in most parts of the world, it can be obtained easily and cheaply [3–5]. Gasoline is a petroleum product, containing alkanes, alkylbenzenes and condensed aromatics [4,8]. While ILs are generally fresh at the moment of delivery to the fire scene, the composition of the IL may change significantly over the course of the fire. Due to temperature and air flow, components of the IL will evaporate. However, due to differences in boiling points of various components within an IL, the extent of weathering is not uniform across all compounds or from one fire scene to the next [4,9,10]. Furthermore, ILs may undergo bacterial degradation if samples are not collected shortly after the fire [4,11,12]. This variability will pose additional challenges for IL detection and identification.

To complicate the problem, debris matrices are also highly variable, often complex, and contain numerous precursor, pyrolysis, and combustion products that interfere with the analysis [4,13]. Investigators will normally select a location that is likely to contain an IL based on evidence such as burn patterns at the scene

\* Corresponding author. Tel.: +1 780 492 8303; fax: +1 780 492 8231.

E-mail addresses: [James.Harynuk@ualberta.ca](mailto:James.Harynuk@ualberta.ca), [jharynuk@gmail.com](mailto:jharynuk@gmail.com) (J.J. Harynuk).

[4,14], or as indicated by aids such as accelerant detection canines [15–17]. Porous materials such as carpet or wood are generally good choices since they are more likely to retain traces of ILs, and being floor coverings, are common substrates to which ILs are delivered [4,18]. Since carpets are made from a variety of natural (e.g. wool, cotton) and synthetic (e.g. polyolefin, nylon, polypropylene) fibers, there is a degree of chemical diversity between different types of carpets. Furthermore, carpets contain dyes, resins, and flame-resistant coatings, and generally have some form of underlay (usually polyurethane foam), which collectively add chemical complexity. Other materials such as paper, plastics, paint, wool, cotton, leather (natural or synthetic), food [4], and even arsonists [19] present at the scene will further complicate the chemical make-up of the matrix.

Matrix components also undergo chemical changes in the fire; temperature and oxygen levels will vary, meaning that a given location in a fire scene may undergo both combustion in the presence of oxygen and pyrolysis in the absence of oxygen over the course of a single fire [4,20,21]. Some of the pyrolysis products generated from the matrix components are also found in ILs; another source of potential confusion when interpreting the data.

The analysis of fire debris involves the concentration of headspace vapors via direct sampling [22], dynamic headspace sampling using sorbent beds [23], passive headspace sampling using activated carbon strips [24,25], or techniques such as solid-phase microextraction (SPME) [26]. Passive headspace extraction (other than by SPME) is typically followed by solvent extraction of the IL residues from the adsorptive medium using a solvent such as CS<sub>2</sub> or occasionally Et<sub>2</sub>O [24,27]. Extracts are then analyzed by gas chromatography mass spectrometry (GC–MS) [4,8].

Once collected, chromatographic data are manually interpreted, typically by two or sometimes three analysts to determine if there are traces of IL present in the debris, and if possible the identity of the IL [8]. This final step is a potentially expensive bottleneck in arson debris analysis that we seek to address through the application of chemometric techniques. The interpretation of the data is a particularly difficult task because of the extreme chemical diversity and complexity of the analytes and matrix, as highlighted above.

## 2. Experimental

All samples were, stored, extracted, and analyzed according to Royal Canadian Mounted Police (RCMP) protocols [28] which follow ASTM methods E1618 and E1412. Briefly, a passive headspace extraction of volatiles onto activated carbon strips (Albrayco Technologies, Cromwell, CT) for 16 h @ 60 °C is performed followed by elution with CS<sub>2</sub> and analysis by GC–MS [8,24]. The only deviation from the standard protocol was the addition of a perdeuterated alkane ladder consisting of n-heptane (d16), n-nonane (d20), n-undecane (d24), n-tridecane (d28), n-pentadecane (d32), n-heptadecane (d36), n-nonadecane (d40) and n-heneicosane (d44) (CDN Isotopes, Pointe-Claire, QC) at concentrations of 16 µL L<sup>-1</sup> each to the solvent (CS<sub>2</sub>) used to elute analytes from the activated carbon strips.

Samples were analyzed using one of three Agilent Technologies 7890A gas chromatographs (GC) with 5975 quadrupole mass spectrometers (MS) and 7683 auto samplers (Agilent Technologies, Mississauga, ON). Data acquisition and automation were accomplished using MS ChemStation (Agilent). The GCs were equipped with 30 m × 250 µm × 0.25 µm HP-1MS columns (Agilent). The oven program used was 40 °C (held for 3.0 min) followed by a ramp to 250 °C at a rate of 8 °C min<sup>-1</sup>, with a final hold of 0.75 min. Samples were injected in split mode into an injector held at 250 °C. Hydrogen carrier gas was used with flow rate of 1.1 mL/min. The injection volume was 1 µL, with a split

ratio of 20:1. The transfer line and source temperatures were 300 and 230 °C, respectively.

Casework samples were analyzed in duplicate at the RCMP laboratory with one sample processed without the perdeuterated ladder for casework, and the second sample processed with the ladder added for this study. The data provided to our laboratory were stripped of all metadata accompanying actual casework samples and given a new set of dummy identifiers. This ensured that no information that could compromise the confidentiality of an investigation was disclosed by the RCMP laboratory.

Chromatograms were exported from Chemstation as comma separated value (.csv) text files and then imported into MATLAB 7.10.0 (The Mathworks, Natick, MA). Chromatograms were aligned on the basis of the deuterated alkane retention ladder [29]. Variable selection to optimize the chemometric models was performed using lab-written backward-elimination/forward selection (BE/FS) hybrid approach [30] guided by two-dimensional cluster resolution (CR) as the model quality metric [29,31]. Final chemometric analysis of the optimized models was performed using lab-written MATLAB routines, and some chemometric analysis functions from the PLS Toolbox 5.2 (Eigenvector Research Inc, Wenatchee, WA). The calculations were performed on an Intel Core i5 750 2.76 GHz processor with 8 GB of RAM and 64-bit Microsoft Windows 7 Professional operating system.

Mass spectral searching to identify selected features in the data set was performed against the NISTMS 2005 library (NIST, Gaithersburg, MD).

## 3. Results and discussion

The time spent by trained scientists to interpret chromatographic data from fire debris is very expensive, as is the time spent training these individuals to a point where they become proficient at the task. A potential solution to this high human cost of data interpretation for arson investigations lies in the development of chemometric models for rapid, objective, and automated identification of ILs in fire debris samples. Should a successful chemometric solution be discovered, it would decrease costs and essentially remove the bottleneck in the analytical procedure. This would increase the overall sample throughput for an arson laboratory while decreasing cost per sample analyzed. This would, by extension, permit fire investigators to increase the number of samples that are taken from a fire scene, and have the results reported more rapidly. As a result, more thorough, faster investigations of fire scenes would be possible.

Previous work has involved the application of exploratory techniques, such as principal components analysis (PCA), to the identification of ILs [5,10,32]. Soft independent modeling of class analogies (SIMCA) has also been used to classify ILs on charred carpet and wood (white pine and poplar) samples [5]. Previously we demonstrated the use of partial least squares-discriminant analysis (PLS-DA) to classify simulated arson debris based on the presence or absence of gasoline [29]. To date, and to the best of our knowledge, there are no reported studies of the successful application of chemometric techniques to the interpretation of actual arson casework samples. Due to extreme conditions and variability of fire scenes, actual casework studies are crucial. The problem with using simulated debris to develop chemometric models is evidenced by the fact that the variables selected to identify gasoline in simulated debris included a series of C<sub>2</sub>-alkylbenzenes [29], which are not generally considered reliable for the identification of gasoline as they can be generated by matrix pyrolysis during the fire [20].

Casework samples used in this study were collected over several months by a variety of arson investigators from fire scenes located across Canada (at the time of the study, the Edmonton

Laboratory handled samples from all jurisdictions in Canada except for Ontario and Quebec). As most arsonists rely on gasoline as the IL, a sufficient number of debris samples could only be obtained for gasoline-containing and gasoline-free debris. Therefore, this initial test on real data focused on the classification of debris based on gasoline content.

With the use of real arson data, there was no control over the contents of the fire scenes, the nature, or amount of ILs being used. The extent of variability in the data was staggering. The amount of gasoline remaining in the debris varied due to differences in the amount of IL used in a given fire, the substrate for the sample, and different extents of combustion and weathering in each fire. Additionally, the composition of gasoline varies depending on factors such as refinery, season, and region of the country [33,34]. The matrix at the fire scenes was completely uncontrolled, and samples were prepared and analyzed by different analysts on one of three GC–MS systems with the same nominal operating conditions. No deviations were made from the standard analytical protocol, with the exception of the addition of the perdeuterated alkane ladder to the desorption solvent for the set of casework samples being directed to this study.

Chemometric techniques are powerful tools for the interpretation of complex analytical data. As a few examples, chemometric techniques have been applied to the determination of fatty acids in cow's milk using FT-IR spectroscopy [35], the analysis of olive oils [36,37], and natural products [38]. Chemometrics have also been previously used in fingerprinting of ignitable liquids and determining their origins during arson investigations [10,29,39,40]. Chromatography, especially when hyphenated to mass spectrometry, provides incredibly rich data. Chemometric techniques can utilize this volume of data to their advantage, and are becoming more popular options for data interpretation. However, chromatographic data requires some sort of chromatographic alignment and variable selection before a chemometric model can be constructed.

### 3.1. Chromatographic alignment

The reason for chromatographic alignment is to ensure that each point along the time axis in each chromatogram is registered in the same variable number in the data matrix being passed to further chemometric analysis. It is required to correct for small shifts in retention time due to slight shifts in temperature, pressure, degradation of stationary phase, and in the case of this study, inter-instrument variations in chromatography.

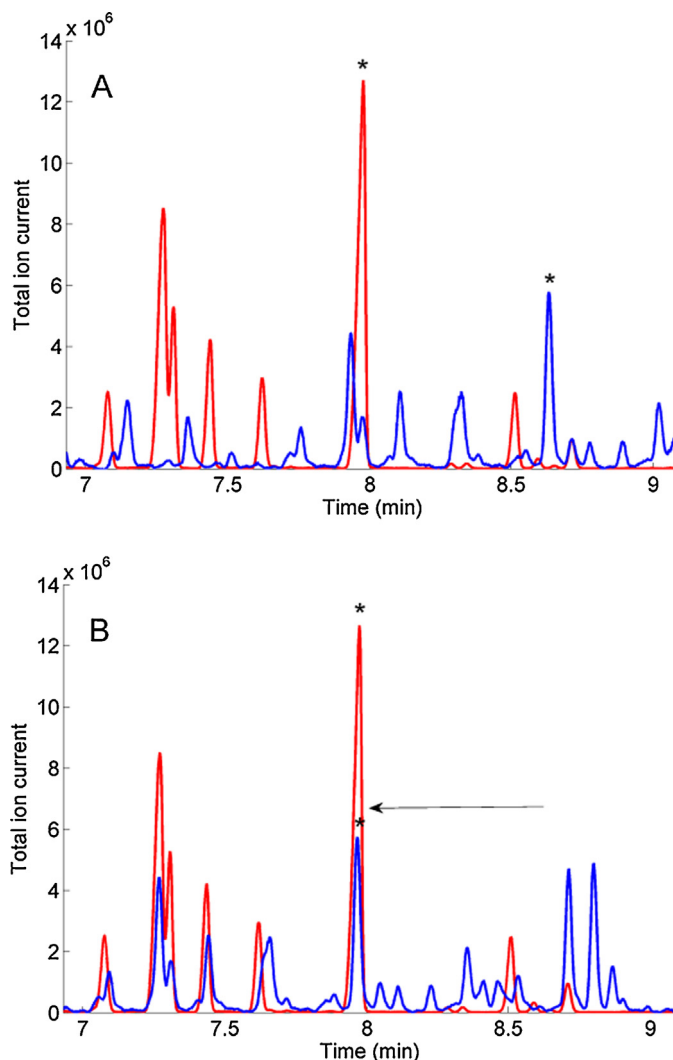
Chromatographic alignment involves comparison of a sample chromatogram with a target chromatogram, followed by shifting and warping of the sample chromatogram's time axis so that its peaks are aligned with those in the target. The target chromatogram can be a chromatogram in the data set [29], or it can be a composite target containing information from multiple chromatograms in the data set [30,31]. Some alignment approaches include piecewise alignment [41], which relies on identification of common peaks in the chromatogram and the alignment target. Other approaches include correlation-optimized warping (COW) [42,43], dynamic time warping (DTW) [43], as well as many others [44–47]. These alignment techniques perform well when the chemical composition of the samples in the dataset remains similar. However, when the chemical profiles of different sample classes (or even samples within a given class) are highly dissimilar, these methods will yield a poor alignment. When the background matrix of the samples is also highly variable, such as the case with arson debris, alignment is even more challenging since the alignment algorithm may be unable to lock onto variables that must be aligned. Arson debris is an example of a challenging alignment problem due to the variability in the nature

of the samples highlighted above. We previously developed a solution to this challenge which relies on the use of a perdeuterated alkane ladder added to the samples [29].

In this study, the perdeuterated alkane ladder method was applied using the product of the extracted ion traces for ions of  $m/z$  34, 50, 66, 80, and 82 as the ladder signal. A randomly selected chromatogram from the training set was chosen as the alignment target. Ion 34, which is due to  $C_2D_5^+$ , was required to add selectivity for some samples of real debris. Due to the use of multiple GC–MS systems to collect the data, extreme shifts in retention times were observed, an example of which is shown in Fig. 1. The perdeuterated alkane ladder approach was able to successfully correct extreme misalignment, even in cases (such as in Fig. 1) where the chromatographic profiles of debris samples were highly dissimilar. Alignment was followed by automated variable selection.

### 3.2. Variable selection

When applying chemometric techniques directly to raw chromatographic data, some form of variable selection is necessary. This is especially true for GC–MS data [48,49]. A variety of variable selection techniques exist, from the simple, such as



**Fig. 1.** Segments of two chromatograms in the casework debris dataset collected on different instruments. A shows unaligned chromatograms. B shows aligned chromatograms. Asterisks indicate a pair of peaks that should be aligned (shift of ~40 s).

integrated peak tables [50,51], single ion [52], or extracted ion [53,54] chromatograms/profiles. To more involved approaches using metrics such as analysis of variance (ANOVA) [29,49,55,56] or selectivity ratio (SR) [57–59] to rank variables followed by selecting a number of top-ranked variables or using a stepwise selection method [60–62]. In this study, relevant variables were selected using SR as a ranking metric followed by a hybrid BE/FS stepwise approach using the cluster resolution (CR) metric [29,30,31] to guide the selection process.

Briefly, CR is a metric which describes the size of the largest confidence ellipses that can be constructed around the two clusters of points that represent a pair of classes of samples plotted in, for example, PCA space, without the ellipses overlapping. Thus it provides a metric that indicates the distance between classes of points while simultaneously accounting for the distribution of points within each class and the relative orientations of the clusters of points. CR is useful for guiding variable selection, but it requires an initial population of variables to be included in the model for it to be calculated. In the BE/FS approach, a population of top-ranked variables most likely to contain important information is chosen and a preliminary model is constructed. This model is evaluated using CR, and then variables are tested individually. This testing is performed by removing the lowest-ranked variable, computing a new candidate model and comparing the CR. If the CR increases in the candidate model, the variable being tested is deemed to be uninformative (or harmful) to the model and it is permanently excluded. If the CR decreases, then the variable is deemed to be important and it is returned to the model. This process continues iteratively testing each variable in turn. After the initial BE, additional variables are checked using an analogous FS approach on the remaining untested variables.

Overall, 156 casework samples were provided by the RCMP. 65 of these samples were confirmed to contain gasoline, 79 were confirmed to contain no IL, and 12 samples were ambiguous. The 156 samples were supplemented by 76 gasoline-free debris samples simulated by the RCMP in accordance with a published protocol [63]. This brought the total number of gasoline-free samples to 155 and the total number of samples for this study to 232.

Each chromatogram consisted of 16,000 scans with  $m/z$  values from 30 to 300, providing a total of 4,336,000 individual variables per chromatogram. Each GC–MS chromatogram was unfolded along the retention time axis, providing a single vector of 4,336,000 variables. For model construction and testing, the data set was separated into three sets: training, optimization, and validation. The 220 chromatograms with known class identities, were randomly split into a training set (110 chromatograms), optimization set (55 chromatograms), and validation set (55 chromatograms). All 12 ambiguous samples were assigned to the validation set, bringing the total number of samples in that set to 67.

During each variable selection step, a candidate 2-component PCA model was constructed using chromatograms from the training set and evaluated using chromatograms from the optimization set. After variable selection, both training and optimization data sets were combined to create the final model which was then tested with chromatograms from the validation set. The initial number of variables used in the BE approach was 10,000 and variables up to rank 25,000 were checked with the FS approach. The flowchart for variable selection is shown in Fig. 2. A total of 1597 variables were selected. Fig. 3 depicts the selected variables and also shows the chemical classes of the compounds that were ascertained by searching the peaks corresponding to the ions against the NISTMS library.

As seen from Fig. 3, C<sub>3</sub>-, C<sub>4</sub>- and C<sub>5</sub>-alkylbenzenes were selected. As mentioned before, gasoline contains light alkanes, alkylbenzenes and condensed aromatics [4,8]. According to standard

method ASTM E 1618, alkanes present in gasoline samples vary by brand, grade and lot. Furthermore, being relatively light molecules, they are more likely to evaporate during gasoline weathering. They are also generated by pyrolysis of some materials (e.g. polyethylene) [64]. Thus one would expect the alkanes to be of little diagnostic value for the purpose of identifying gasoline in arson debris. This explains the exclusion of light alkanes by the algorithm.

ASTM E 1618 also cautions against using BTEX (benzene, toluene, ethylbenzene, xylenes) and condensed aromatics such as naphthalene as markers for gasoline. These compounds are also natively present even in a gasoline-free debris matrix as they can be formed by numerous pyrolysis processes [20]. It is reassuring that the automated approach to variable selection also ignored this group of compounds. ASTM E 1618 recommends using the C<sub>3</sub>-, C<sub>4</sub>-, and C<sub>5</sub>-alkylbenzenes as markers for gasoline as these compounds are characteristic of gasoline and do not generally have other sources in debris. As seen in Fig. 3, variable selection guided by the CR metric selected variables originating from the compounds recommended by the standard method for identifying gasoline in fire debris. It is important to note that the selection was performed automatically without any direction as to which variables to focus on. In fact the only information provided to the algorithm was the binary class assignment (gasoline/no gasoline) of the chromatograms.

### 3.3. Model construction

Following selection of relevant variables, chemometric models for classification of arson debris were constructed. Pre-processing for all chemometric models involved normalization of each signal to an area of 1 followed by autoscaling of the combined training and optimization sets used to construct the final model. The autoscaling parameters determined in this step were then applied to the validation data.

Initially, a PLS-DA classification model was constructed (Fig. 4). The number of latent variables (LVs) was chosen using venetian blinds cross-validation [65] with 10 data splits and using the number of LVs that provided the lowest misclassification rate. Three LVs were used in the model construction.

As seen from Fig. 4, the PLS-DA model correctly classified all samples in the gasoline-containing and gasoline-free groups. Briefly, PLS-DA (as implemented here) must classify a sample as being either a member of Class A (gasoline) or Class B (not gasoline). The selected variables are projected through the model and a predicted  $y$ -value is calculated. If this value is greater than some threshold (as shown by the dashed line in Fig. 4), then the sample is deemed to be the member of Class A, and if it is below, it is deemed to be a member of Class B. Some of the ambiguous samples projected into the gasoline-containing class while many remained near the classification border. However, PLS-DA is likely not the most appropriate technique for gasoline classification. The reason is that PLS-DA assigns a value of zero for all samples in the gasoline free-class. However, assigning the same  $y$ -value to all the samples in the gasoline-free class is not sensible: the only similarity between samples in the gasoline-free class is the lack of gasoline. The chemical composition of one non-gasoline containing sample can be completely different from the chemical composition of another sample in the same class.

To address this issue, SIMCA was tested as a modeling tool. SIMCA differs from PLS-DA in that it does not force a yes/no decision on a sample. Instead SIMCA creates a PCA model for one or more selected classes or groups of classes [48]. The samples are then projected into the collection of PCA models. Class assignment is made on the basis of residual scores: as residual scores for a sample in a class model increase, the likelihood of class



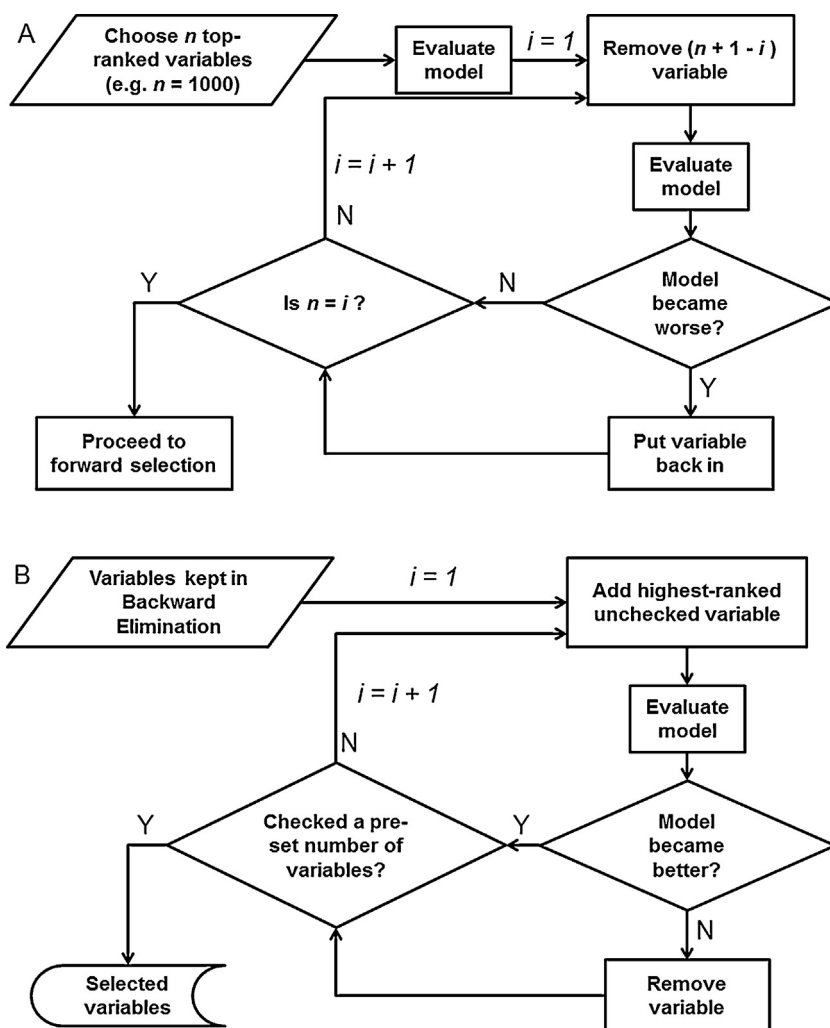


Fig. 2. Variable selection techniques used. (A) Backwards elimination; (B) forward selection. CR metric used in the evaluate model step.

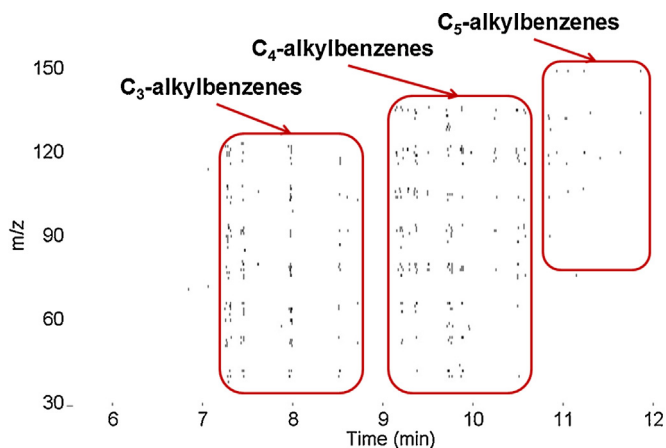


Fig. 3. Variables from GC-MS chromatograms included in optimized model for identification of gasoline in arson debris. Black dots represent variables used in model construction.

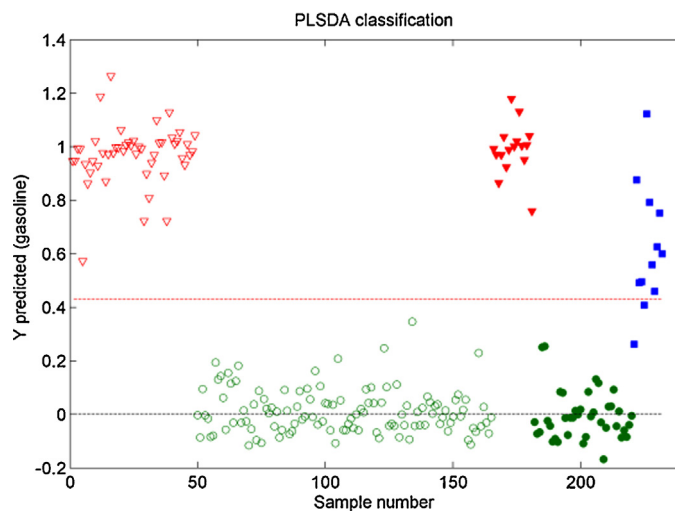
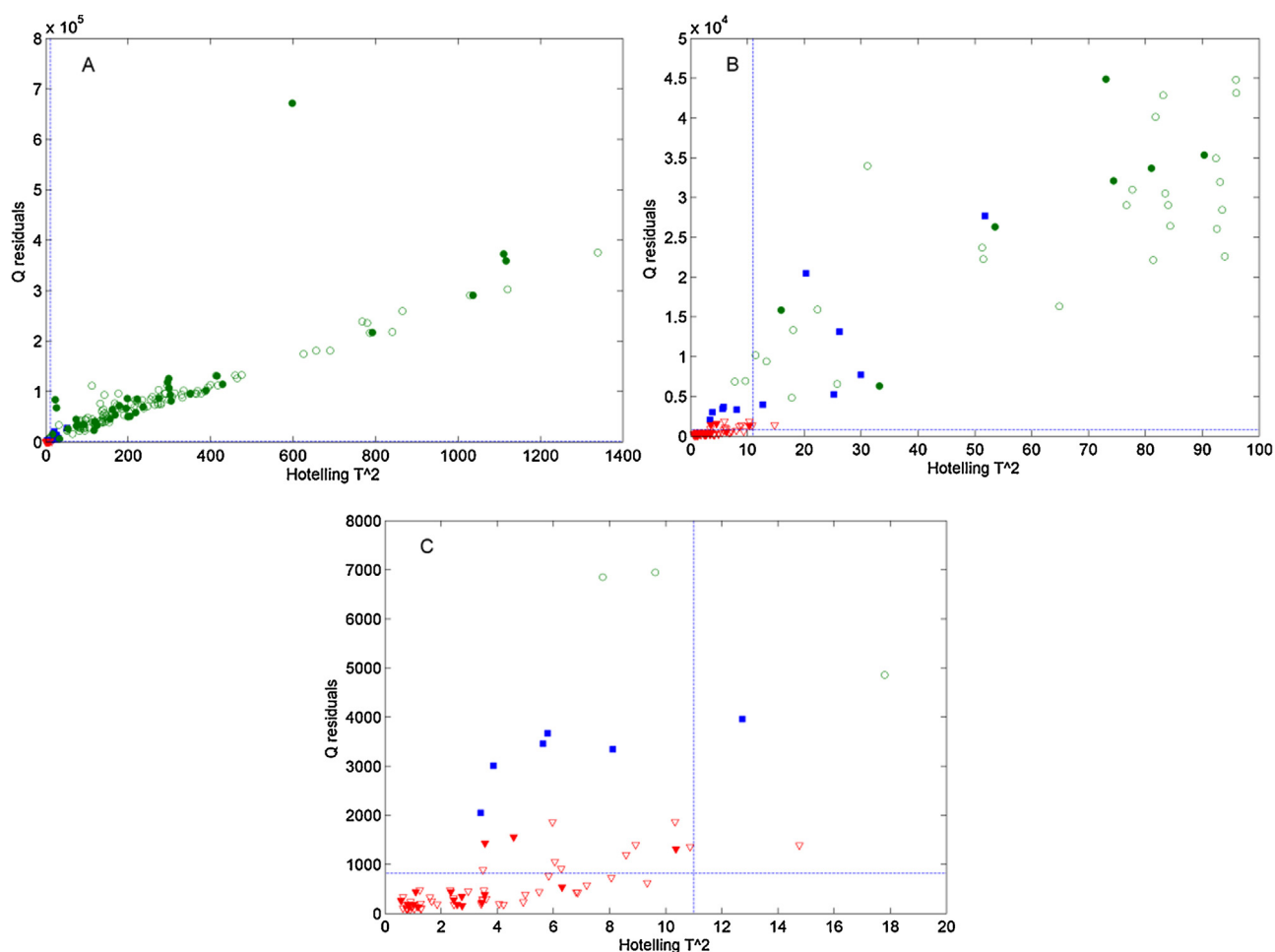


Fig. 4. PLS-DA plot for arson data. Red triangles indicate gasoline-containing samples. Green circles indicate gasoline-free samples. Blue squares indicate ambiguous samples. Hollow markers indicate training and optimization set data. Filled markers indicate validation set data. (For interpretation of the references to color in this text, the reader is referred to the web version of the article.)



**Fig. 5.** SIMCA plot for arson data. Red triangles indicate gasoline-containing samples. Green circles indicate gasoline-free samples. Blue squares indicate ambiguous samples. Hollow markers indicate training and optimization set data. Dashed lines indicate 95% confidence levels for Hotelling  $T^2$  and Q residuals. Filled markers indicate validation set data. A, B and C show different zoom levels for the plot. (For interpretation of the references to color in this text, the reader is referred to the web version of the article.)

membership for the sample in the particular class decreases. Unlike PLS-DA, SIMCA allows a sample to be a member of none, one, or multiple classes. In the case of fire debris, it is possible that a mixture of ILs was used, making application of a technique that allows multiple class membership more appropriate.

The number of PCs for the gasoline SIMCA model was chosen using venetian blinds cross-validation with 10 data splits. Four PCs provided the lowest error of cross-validation so they were chosen for the final model construction. Considering residuals for the gasoline model, if a sample has a pattern in the selected variables that is similar to gasoline, then it will have very low values for its Q and Hotelling  $T^2$  residuals. On the other hand, if a sample contains no gasoline, it will not fit the gasoline model well and it will have high residual values. The Q vs. Hotelling  $T^2$  plot for the gasoline data set is presented in Fig. 5 at several axis magnifications. Gasoline-containing samples should lie in the bottom left corner of this plot, and as samples become less gasoline-like, they should drift toward the top right corner of the plot, as observed.

Comparing the results in Fig. 5 to those in Fig. 4, SIMCA was also able to reliably classify arson samples based on gasoline content, including the ambiguous samples that were not classified by PLS-DA.

#### 4. Conclusions

Deuterated alkane ladder-based alignment and a CR-guided automated approach to variable selection have been applied to

generate PLS-DA and SIMCA models for the classification of casework arson debris samples on the basis of gasoline content. The alignment was able to account for extreme retention time shifts ( $\sim 40$  s). The variable selection algorithm automatically selected a suite of variables derived from compounds identified in the standard ASTM method as being reliable markers for gasoline, while successfully ignoring compounds known to be unreliable markers of gasoline. The final PLS-DA and SIMCA models were able to reliably classify the samples as being either gasoline-containing or gasoline-free, with no false positives or false negatives. To the best of our knowledge, this is the first demonstration of the successful application of chemometric classification techniques to casework arson data.

While more research and long-term testing of the models developed in this work are required, as well as testing on data from other laboratories, this work represents a significant step toward the development of an expert system for interpreting arson data. Such a system could be imagined to be similar to those already routinely used to aid in DNA analysis, acting as a “second scientist” to review data [66,67].

#### Acknowledgements

The authors would like to acknowledge the Department of Chemistry at the University of Alberta, and the Natural Sciences and Engineering Research Council (NSERC) Canada for their support of this research. Alberta Innovates Technology Futures’

support through a New Faculty Award to Dr. Harynuk is also acknowledged.

## References

- [1] Arson Oxford English Dictionary [Online]; Oxford University Press, Second edition, 1989; online version June 2012, <http://www.oed.com/view/Entry/11118> (accessed June 18, 2012).
- [2] D. Roy, The Cost of Fires: a Review of the Information Available, Home Office, London, 1997.
- [3] P.M.L. Sandercock, Fire investigation and ignitable liquid residue analysis – a review: 2001–2007, *Forensic Sci. Int.* 176 (2008) 93–110.
- [4] E. Stauffer, J.A. Dolan, R. Newman, *Fire Debris Analysis*, Elsevier, Inc., Amsterdam, 2008.
- [5] B. Tan, J.K. Hardy, R.E. Snavely, Accelerant classification by gas chromatography/mass spectrometry and multivariate pattern recognition, *Anal. Chim. Acta* 422 (2000) 37–46.
- [6] J.P. Jackowski, The incidence of ignitable liquid residues in fire debris as determined by a sensitive and comprehensive analytical scheme, *J. Forensic Sci.* 42 (1997) 828–832.
- [7] D.A. Sutherland, Fire debris analysis statistics and the use of the latest analytical tools, *J. Can. Assoc. Fire Invest.* (1999) 11–13.
- [8] ASTM Standard E1618, 2006.1, Standard Test Method for Ignitable Liquid Residues in Extracts from Fire Debris Samples by Gas Chromatography–Mass Spectrometry, ASTM International, West Conshohocken, PA, 2006, <http://dx.doi.org/10.1520/E1618-11> www.astm.org.
- [9] B.M. Zorzetti, J.M. Shaver, J.J. Harynuk, Estimation of the age of a weathered mixture of volatile organic compounds, *Anal. Chim. Acta* 694 (2011) 31–37.
- [10] P.M.L. Sandercock, E. Du Pasquier, Chemical fingerprinting of gasoline: 2. Comparison of unevaporated and evaporated automotive gasoline samples, *Forensic Sci. Int.* 140 (2004) 43–59.
- [11] D.C. Mann, W.R. Gresham, Microbial degradation of gasoline in soil, *J. Forensic Sci.* 35 (1990) 913–923.
- [12] K.P. Kirkbride, S.M. Yap, S. Andrews, P.E. Pigou, G. Klass, A.C. Dinan, F.L. Peddie, Microbial degradation of petroleum hydrocarbons: implications for arson residue analysis, *J. Forensic Sci.* 37 (1992) 1585–1599.
- [13] E.S. Bodle, J.K. Hardy, Multivariate pattern recognition of petroleum-based accelerants by solid-phase microextraction gas chromatography with flame ionization detection, *Anal. Chim. Acta* 589 (2007) 247–254.
- [14] J.F. O'Donnell, The sampling of burned areas for accelerant residues analysis, *Fire Arson Investig.* 35 (1985) 18–20.
- [15] M.E. Kurz, M. Billard, M. Rettig, J. Augustiniak, J. Lange, M. Larsen, R. Warrick, T. Mohns, R. Bora, K. Broadus, G. Hartke, B. Glover, D. Tankersley, J. Marcouiller, Evaluation of canines for accelerant detection at fire scenes, *J. Forensic Sci.* 39 (1994) 1528–1536.
- [16] M.E. Kurz, S. Schultz, J. Griffith, K. Broadus, J. Sparks, G. Dabdoub, J. Brock, Effect of background interference on accelerant detection by canines, *J. Forensic Sci.* 41 (1996) 868–873.
- [17] S.R. Katz, C.R. Midkiff, Unconfirmed canine accelerant detection: a reliability issue in court, *J. Forensic Sci.* 43 (1998) 329–333.
- [18] W. Bertsch, Q.-W. Zhang, Sample preparation for the chemical analysis of debris in suspect arson cases, *Anal. Chim. Acta* 236 (1990) 183–195.
- [19] 1999 Darwin Award: Firefighters Ignite! 2012 (accessed 08.11.12) <http://www.darwinawards.com/darwin/darwin1999-24.html>.
- [20] E. Stauffer, Basic concept of pyrolysis for fire debris analysts, *Sci. Justice* 43 (2003) 29–40.
- [21] J.D. DeHaan, D.J. Brien, R. Large, Volatile organic compounds from the combustion of human and animal tissue, *Sci. Justice* 44 (2004) 223–236.
- [22] ASTM Standard E1388, Standard Practice for Sampling of Headspace Vapors from Fire Debris Samples, ASTM International, West Conshohocken, PA, 2005, <http://dx.doi.org/10.1520/E1388-05> www.astm.org.
- [23] ASTM Standard E1413, Separation and Concentration of Ignitable Liquid Residues from Fire Debris Samples by Dynamic Headspace Concentration, ASTM International, West Conshohocken, PA, 2007, <http://dx.doi.org/10.1520/E1413-07> www.astm.org.
- [24] ASTM Standard E1412, Standard Practice for Separation of Ignitable Liquid Residues from Fire Debris Samples by Passive Headspace Concentration With Activated Charcoal, ASTM International, West Conshohocken, PA, 2007, <http://dx.doi.org/10.1520/E1412-07> www.astm.org.
- [25] K. Cavanagh, E. Du Pasquiere, C. Lennard, Background interference from car carpets – the evidential value of petrol residues in cases of suspected vehicle arson, *Forensic Sci. Int.* 125 (2002) 22–36.
- [26] K.G. Furton, J.R. Almirall, M. Bi, J. Wang, L. Wu, Application of solid-phase microextraction to the recovery of explosives and ignitable liquid residues from forensic specimens, *J. Chromatogr. A* 885 (2000) 419–432.
- [27] D. Massey, E. Du Pasquier, C. Lennard, Solvent desorption of charcoal strips (DFLEX) in the analysis of fire debris samples: replacement of carbon disulfide, *Can. Soc. Forensic Sci. J.* 35 (2002) 195–207.
- [28] M. Nowlan, A.W. Stuart, G.J. Basara, P.M.L. Sandercock, Use of a solid absorbent and an accelerant detection canine for the detection of ignitable liquids burned in a structure fire, *J. Forensic Sci.* 52 (2007) 643–648.
- [29] N.A. Sinkov, B.M. Johnston, P.M.L. Sandercock, J.J. Harynuk, Automated optimization and construction of chemometric models based on highly variable raw chromatographic data, *Anal. Chim. Acta* 697 (2011) 8–15.
- [30] N.A. Sinkov, J.J. Harynuk, Three-dimensional cluster resolution for guiding automatic chemometric model optimization, *Talanta* 103 (2013) 252–259.
- [31] N.A. Sinkov, J.J. Harynuk, Cluster resolution: a metric for automated, objective and optimized feature selection in chemometric modeling, *Talanta* 83 (2011) 1079–1087.
- [32] P. Doble, M. Sandercock, E. Du Pasquier, P. Petocz, C. Roux, M. Dawson, Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks, *Forensic Sci. Int.* 132 (2003) 26–39.
- [33] P.M.L. Sandercock, A survey of Canadian gasolines (2004), *Can. Soc. Forensic Sci. J.* 40 (2007) 105–130.
- [34] P.M.L. Sandercock, Survey of Canadian Gasoline (Winter 2010), *Can. Soc. Forensic Sci. J.* 45 (2012) 64–78.
- [35] M. Ferrand, B. Huquet, S. Barbey, S. Barillet, F. Faucon, H. Larroque, O. Leray, J.M. Trommenschlager, M. Brochard, Determination of fatty acid profile in cow's milk using mid-infrared spectrometry: interest of applying a variable selection by genetic algorithms before a PLS regression, *Chemom. Intell. Lab. Syst.* 106 (2011) 183–189.
- [36] P. de la Mata-Espinosa, J.M. Bosque-Sendra, R. Bro, L. Cuadros-Rodríguez, Olive oil quantification of edible vegetable oil blends using triacylglycerols chromatographic fingerprints and chemometric tools, *Talanta* 85 (2011) 177–182.
- [37] P. de la Mata, A. Dominguez-Vidal, J.M. Bosque-Sendra, A. Ruiz-Medina, L. Cuadros-Rodríguez, M.J. Ayora-Cañada, Olive oil assessment in edible oil blends by means of ATR-FTIR and chemometrics, *Food Control* 23 (2012) 449–455.
- [38] O.M. Kvalheim, H.Y. Chan, I.F.F. Benzie, Y.T. Szeto, A.H.C. Tzang, D.K.W. Mok, F.T. Chau, Chromatographic profiling and multivariate analysis for screening and quantifying the contributions from individual components to the bioactive signature in natural products, *Chemom. Intell. Lab. Syst.* 107 (2011) 98–105.
- [39] P.M.L. Sandercock, E. Du Pasquier, Chemical fingerprinting of unevaporated automotive gasoline samples, *Forensic Sci. Int.* 134 (2003) 1–10.
- [40] P.M.L. Sandercock, E. Du Pasquier, Chemical fingerprinting of gasoline: Part 3. Comparison of unevaporated automotive gasoline samples from Australia and New Zealand, *Forensic Sci. Int.* 140 (2004) 71–77.
- [41] K.J. Johnson, B.W. Wright, K.H. Jarman, R.E. Synovec, High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis, *J. Chromatogr. A* 996 (2003) 141–155.
- [42] N.-P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping, *J. Chromatogr. A* 805 (1998) 17–35.
- [43] G. Tomasi, F. van den Berg, C. Andersson, Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data, *J. Chemom.* 18 (2004) 231–241.
- [44] P.H.C. Eilers, Parametric time warping, *Anal. Chem.* 76 (2004) 404–411.
- [45] S. Toppo, A. Roveri, M.P. Vitale, M. Zaccarin, E. Serain, E. Apostolidis, M. Gion, M. Mariolino, F. Ursini, MPA: a multiple peak alignment algorithm to perform multiple comparisons of liquid-phase proteomic profiles, *Proteomics* 8 (2008) 250–253.
- [46] A.M. Van Norderkassel, M. Dzykowski, P.H.C. Eilers, Y. Vander Heyden, A comparison of three algorithms for chromatograms alignment, *J. Chromatogr. A* 1118 (2006) 199–210.
- [47] W. Yao, X. Yin, Y. Hu, A new algorithm of piecewise automated beam search for peak alignment of chromatographic fingerprints, *J. Chromatogr. A* 1160 (2007) 254–262.
- [48] R.G. Brereton, *Chemometrics Data Analysis for the Laboratory and Chemical Plant*, Wiley, Toronto, 2003.
- [49] K.J. Johnson, R.E. Synovec, Pattern recognition of jet fuels: comprehensive GC × GC with ANOVA-based feature selection and principal component analysis, *Chemom. Intell. Lab. Syst.* 60 (2002) 225–237.
- [50] B.T. Weldegergis, A.M. Crouch, Analysis of volatiles in Pinotage wines by stir bar sorptive extraction and chemometric profiling, *J. Agric. Food Chem.* 56 (2008) 10225–10236.
- [51] R.B. Gaines, G.J. Hall, G.S. Frysinger, W.R. Gronlund, K.L. Juare, Chemometric determination of target compounds used to fingerprint unweathered diesel fuels, *Environ. Forens.* 7 (2006) 77–87.
- [52] J.H. Christensen, G. Tomasi, Practical aspects of chemometrics for oil spill fingerprinting, *J. Chromatogr. A* 1169 (2007) 1–22.
- [53] C.R. Borges, Concept for facilitating analyst-mediated interpretation of qualitative chromatographic-mass spectral data: an alternative to manual examination of extracted ion chromatograms, *Anal. Chem.* 79 (2007) 4805–4813.
- [54] L.J. Marshall, J.W. McIlroy, V.L. McGuffin, R.W. Smith, Association and discrimination of diesel fuels using chemometric procedures, *Anal. Bioanal. Chem.* 394 (2009) 2049–2059.
- [55] N.E. Watson, M.M. VanWingerden, K.M. Pierce, B.W. Wright, R.E. Synovec, Classification of high-speed gas chromatography–mass spectrometry data by principal component analysis coupled with piecewise alignment and feature selection, *J. Chromatogr. A* 1129 (2006) 111–118.
- [56] K.M. Pierce, J.K. Hope, K.J. Johnson, B.W. Wright, R.E. Synovec, Classification of gasoline data obtained by gas chromatography using a piecewise alignment algorithm combined with feature selection and principal component analysis, *J. Chromatogr. A* 1096 (2005) 101–110.
- [57] T. Rajalahti, R. Arneberg, F.S. Berven, K.M. Myhr, R.J. Ulvik, O.M. Kvalheim, Biomarker discovery in mass spectral profiles by means of selectivity ratio plot, *Chemom. Intell. Lab. Syst.* 95 (2009) 35–48.

- [58] T. Rajalahti, R. Arneberg, A.C. Kroksveen, M. Berle, K.M. Myhr, O.M. Kvalheim, Discriminating variable test and selectivity ratio plot: quantitative tools for interpretation and variable (biomarker) selection in complex spectral or chromatographic profiles, *Anal. Chem.* 81 (2009) 2581–2590.
- [59] J.P.M. Andries, Y. Vander Heyden, L.M.C. Buydens, Improved variable reduction in partial least squares modelling based on predictive-property-ranked variables and adaptation of partial least squares complexity, *Anal. Chim. Acta* 705 (2011) 292–305.
- [60] I. Guyon, A.J. Elisseeff, An introduction to variable and feature selection, *J. Mach. Learn. Res.* 3 (2003) 1157–1182.
- [61] R.R. Hocking, The analysis and selection of variables in linear regression, *Biometrics* 32 (1976) 1–49.
- [62] D.E. Axelson, *Data Preprocessing for Chemometric and Metabolomic Analysis*, first ed., MRi Consulting, Kingston, ON, 2010.
- [63] P.M.L. Sandercock, Preparation of pyrolysis reference samples: evaluation of a standard method using a tube furnace, *J. Forensic Sci.* 57 (2012) 738–743.
- [64] J.A. Onwudili, N. Insura, P.T. Williams, Composition of products from the pyrolysis of polyethylene and polystyrene in a closed batch reactor: Effects of temperature and residence time, *J. Anal. Appl. Pyrolysis* 86 (2009) 293–303.
- [65] A.D. Walmsley, Chemometrics and data treatment, in: J.M. Chalmers (Ed.), *Spectroscopy in Process Analysis*, Sheffield Academic Press, Sheffield, England, 2000 (Chapter 10).
- [66] M.R. Aniba, J.D. Thompson, Knowledge based expert systems in bioinformatics, in: P. Vizureanu (Ed.), *Expert Systems. InTech*, 2010, p. 185.
- [67] M. Bill, P. Gill, J. Curran, T. Clayton, R. Pinchin, M. Healy, J. Buckleton, PENDULUM – a guideline-based approach to the interpretation of STR mixtures, *Forensic Sci. Int.* 148 (2005) 181–189.